Running Head: Standards for Educational Research

Standards in Conducting and Publishing Research in Education

Bruce Thompson

Texas A&M University
and
Baylor College of Medicine
http://www.coe.tamu.edu/~bthompson

Bruce Thompson is Distinguished Professor of Educational Psychology and CEHD Distinguished Research Fellow, and Distinguished Professor of Library Sciences, Texas A&M University, and Adjunct Professor of Allied Health Sciences, Baylor College of Medicine (Houston). He is past editor of Educational and Psychological Measurement, the series, Advances in Social Science Methodology, and two other journals. He is the author/editor of nearly 200 articles, and several books, including the recently published Foundations of Behavioral Statistics. His contributions have been especially influential in moving the field as regards greater emphasis on effect size reporting and interpretation, and promoting improved understanding of score reliability. *Bruce is (a) co-editor of the teaching, learning, and human development section of the American Educational Research Journal (AERJ:TLHD), (b) a former member of AERA Council, and (c) a previous nominee for AERA President.*

Abstract

Arthur Levine's (2007) recent report on the academic preparation of education doctoral students, <u>Educating Researchers</u>,

http://www.edschools.org/EducatingResearchers/educating_researchers.pdf

is highly critical of both AERA and the scholarly training of education doctoral students. One criticism Levine (2007) offers is his view that educational researchers have no standards with which to evaluate their own and each others' work. The report comes on the heals of Mosteller and Boruch's (2002) <u>Evidence Matters</u>, and Towne and Shavelson's (2003) <u>Scientific Research in Education</u> (cf. Towne, Wise & Winters, 2005). These reports provide context for discussing AERA's (2006) "Standards for Reporting on Empirical Social Science Research in AERA Publications,"

http://www.aera.net/uploadedFiles/Opportunities/
StandardsforReportingEmpiricalSocialScience_PDF.pdf

and related developments, such as the 2005 creation of the Society for Research on Educational Effectiveness.

Standards in Conducting and Publishing Research in Education

In 1975, as a masters' student I attended my first AERA annual meeting and presented my first conference paper. I felt I had never encountered so many bright and attractive people gathered in a single location. I thought I had died and gone to heaven.

I especially treasured the copies of the myriad presentations distributed at the annual meeting. In my early experiences at AERA, virtually all the sessions included discussants who had faithfully been provided advance copies of the papers, so that discussant remarks could be thoughtfully prepared. And presenters actually wrote papers, and routinely made the papers available to audience members.

I have been to every annual meeting since then. But lately I find myself thinking, as I pack my suitcases and papers each year, why am I doing this? What is AERA today about, anyway? Maybe I have changed, and simply reached a more contemplative stage of life. But the annual meeting too has certainly changed a lot over the years, and so has AERA itself.

Some AERA Past Presidents' Views

In 1998, in the note on annual meetings typically published in the June/July issues of Educational Researcher by AERA Program Chairs and AERA Presidents, as regards the 1999 meeting Geoffrey Saxe and Alan Schoenfeld quoted from a letter that they had recently received:

> At [annual meetings] we had a hard time finding rigorous research that reported actual conclusions.

Perhaps we should rename the association the American Educational Discussion Association... (p. 41)

Saxe and Schoenfeld (1998) then suggested, "let us recall that at AERA, 'research' is our middle name!" (p. 41). I felt some sympathy with the views in the letter they quoted. And I was disquieted that an AERA President and Program Chair deemed it necessary to remind members about the importance of research within AERA.

In 2003, a presidential invited session featured several AERA past presidents' reflections on the state of educational research and the AERA culture. The audio of these comments (session #45.010) is available at

<http://www.softconference.com/storefront/230421>.

Past president Penelope Peterson noted that:

Would things be better or worse off if we didn't have AERA? I think AERA is a microcosm of these very problems. In what ways? Well problems with quality. Like the field of education research, the AERA meeting is made up of sessions that vary greatly in quality and often tend to be on the low end... I don't think it's possible to change AERA. I concluded this after being president in 1997. We tried very hard, a number of us tried to change things. I don't think we changed much. I don't think we succeeded. After I was president in 1997, I concluded what we really needed was to create a new and different organization, and that's what I think

we need to do today. I'm even more convinced of it today.

Past president Jim Popham argued,

[I]t seems to me that part of our problems stems directly from the name of our Association. The two key words are educational and research, and... "research" is not the middle. You can't have a single middle word in a four-word title... "Educational" and "research" are the two ones that I want to focus on. I've been a member for 45 years now and during that time, dues paid promptly on every occasion, during that time it seems to me that research is winning decisively over educational, and I think that has to be shifted... My one point is that we need new evaluative criteria, by which to judge the quality of our work, and those criteria have to relate very directly to education... I do think that if we do not improve the educational process, we shouldn't be in business.

And past president Alan Schoenfeld observed,

[E]very theory proves absolutely everything... Where does this get us? It gets us into stupid arguments in the pages of Educational Researcher where a theorist goes my theory is better than your theory, which makes us look like a bunch of imbeciles to the world at large... My bottom line is that we're at the turning point. Yes, it's true there are people

who are out to get us, and a bunch of them are morons who wouldn't recognize science if it jumped up and bit them on the butt. But the fact of the matter is that we've contributed to that, because we haven't gotten our act together in the right way. And we can begin to fix that by being much more serious about the research end of things, and much more serious about the connections between research and practice... We do it right and we can begin to do it properly, but it's going to take a serious effort on our part, and serious culture change.

In 2004, in a different venue, AERA past president Gene Glass offered some similar views. He noted that, "AERA has more to do with legitimizing certain messages... than it does with advancing our understanding of education" (Robinson, 2004, p. 29).

Reflections of Paradigm Conflicts, or Not

Some readers of the Saxe and Schoenfeld (1998) note may have dismissed these concerns as merely reflecting a paradigm war between quantitative and qualitative researchers. But I think AERA on the whole is past that (Larabie, 2003), and I feel I am as well (Capraro & Thompson, in press).

At a recent meeting at her house, with a student whose mixed-methods dissertation we were co-chairing, Yvonna Lincoln declared that I am "a quantitative researcher with a heart!" (This declaration may have been facilitated by the cabernet that the three of us were enjoying at the time.)

Certainly neither quantitative nor qualitative researchers are

monolithic in their views. Researchers of all stripes occasionally do stupid things (Lagemann, 2000). For example, as regards poor qualitative work, Fetterman (1982) noted that, "In one study, labelled 'An ethnographic study of...,' observers were on site only one point in time for 5 days" (p. 17).

My favorite example of dumb and dumber of the quantitative ilk involves Eisner's (1983) study of articles published in two volumes of the American Educational Research Journal:

> [Treatments are often] so brief that the achievement of educationally significant results is highly unlikely. The median experimental time for seven of the 15 experimental studies that reported experimental treatment time in Volume 18 of the AERJ is 1 hour and 15 minutes. I suppose that we should take some comfort in the fact that this represents a 66 percent increase over a 3-year period. In 1978 the median experimental treatment time per subject was 45 minutes. (p. 14)

But no thoughtful quantitative researcher who thinks of names such as Piaget or Freud can blithely dismiss the seminal contributions that qualitative work, even sometimes qualitative work with small numbers of participants, can make. And when qualitative and quantitative methods are used together, sometimes beautiful things can happen in a given project or series of studies (cf. Cook & Heath, 2001; Thompson, Kyrillidou & Cook, in press).

Related Disciplines

I think at least some of my disquiet arises out of my

awareness that other disciplines have taken various steps to articulate visions of one of more forms that improved intellectual inquiry might take. These may prove ultimately to be missteps, but there seems to be reflection leading to some argument being presented about getting better, or at least what various forms of good might look like. Fiona Fidler (2005) has reviewed the movements of various fields in her amazing doctoral dissertation, "From statistical significance to effect estimation: Statistical reform in psychology, medicine, and ecology."

For example, in psychology, after two years of debate regarding whether statistical significance tests should be banned from journals of the American Psychological Association (APA), the APA Board of Scientific Affairs appointed its Task Force on Statistical Inference (Shea, 1996). The Task Force did not endorse a ban on these tests, but did articulate a wide-ranging set of recommendations for improved inquiry and reporting (Wilkinson & APA Task Force on Statistical Inference, 1999), some of which influenced the admonitions in the recent fifth edition of the APA (2001) <u>Publication Manual</u> (cf. p. 22, pp. 25-26). However, as the Australian, Fiona Fidler (2002), observed, "Of the major American associations, only [the editorial policies of] all the journals of the American Educational Research Association have remained silent on all these issues" (p. 754).

In medicine, journal editors from around the world got together and articulated a uniform set of minimal manuscript requirements that they thought would improve research practice (International Committee of Medical Journal Editors, 1997).

Notwithstanding some progress in psychology, the movement within medicine caused psychologists Finch, Cumming and Thomason (2001) to lament, at least as regards quantitative inquiry,

> Why has reform proceeded further in some other disciplines, including medicine, than in psychology? ...What has happened in psychology was not inevitable. We leave to historians and sociologists of science the fascinating and important question of why psychology has persisted for so long with poor statistical practice. (pp. 205-206)

Where is AERA at?

Of course, AERA has not been completely silent regarding concerns for research rigor. For example, in 1998 the AERA Research Advisory Committee was created. The Research Committee is charged with:

> (1) Fostering high-quality research and discussions thereof; (2) the preparation of new researchers; and (3) funding and other infrastructure issues. The committee... [is intended] to strengthen the research-related capacity of AERA and its members, coordinate its activities with appropriate AERA programs, and be entrepreneurial in nature. (American Educational Research Association, 2002, p. 1)

And AERA has one among the 16 or so organizational representatives on the Joint Committee for Standards in Educational Evaluation. In 1994 the Joint Committee published the program

evaluation standards (Sanders & Joint Committee, 1994), which were the first standards for professional conduct ever certified as the American standards (Sanders, 1994) by the American National Standards Institute (ANSI). The Joint Committee has also published the student evaluation standards.

But as I reflect on recent perusals of AERA annual meeting programs and AERA journals, I am left with a sense of disquiet that AERA has developed a *laissez faire* culture, or perhaps a counter-culture, which actively seeks to avoid making any judgments regarding research rigor or value. Indeed, as suggested by the letter quoted by Saxe and Schoenfeld (1998), it is not clear that we have even general agreement regarding what is research and what it is not, or what, if any, boundaries distinguish inquiry from opinion, or what criteria, if any, might be used to evaluate or interpret research.

My concern is not that we seemed to have evolved into a *laissez faire* culture. There may be good reasons to prefer such a climate. Rationale involving academic freedom and a general preference for openness as against proscription might be cited.

But somehow we are selecting 50% of submitted proposals for presentation at the annual meeting, and 15% of submitted manuscripts for publication. Are these decisions being governed solely by the vagaries of a given editorial panel or the chance selection of a particular set of referees?

My disquiet is that I believe educational research can and should make a difference in the lives of children in classrooms, clients in therapy, and so forth, just as medical research (Sackett, Straus, Richardon, Rosenberg & Haynes, 2000) and

psychological research (Chambless, 1998) can make real differences in people's lives. Of course, calls for evidence-based practice in education have also been promulgated (Mosteller & Boruch, 2002). For me research isn't just a game to play to secure tenure, or only a vehicle for securing academic fame (though almost never fortune). I think we have to take educational research seriously, because I think education is serious stuff.

I do not rule out the fact that some poor studies will always be done. And certainly my personal view is that even well-done work has value mainly in the form of one contribution to the corpus of studies, and not as a single study (Thompson, 2002, 2006).

But I feel as though we have ended up in this place without having had a real conversation about where we wanted to be as an organization. For example, I do not buy an argument that standards and rigor are irrelevant as regards qualitative research (cf. LeCompte & Goetz, 1982). Nor do I believe that arguments regarding paradigm conflicts are the same as arguments about what AERA is or should be.

It's not the place we are at that I regret. It's the conversation that I feel we haven't had that I miss. I would very much like for us to have that conversation, and see where it takes us, even if we end up in a place much like the one where we started. Even if we are now the American Educational Discussion Association, or the American Educational Opinion Association, we ought to be able to have a thoughtful dialogue about who we are. I think educational research is a serious business, with potentially important impacts on educational stakeholders, and if it is, I think such a conversation is important as well.

It's Not "the Economy, Stupid," It's the Culture!

In their article in the Educational Researcher, Feuer, Towne and Shavelson (2002b) asked,

> What are the most effective means of stimulating more and better scientific educational research? ...[T]he primary emphasis [italics added] should be on nurturing and reinforcing a scientific culture of educational research... [T]he development of a scientific culture rests with individual researchers, supported by leadership in their professional associations..." (p. 4)

They defined scientific culture as "a set of norms and practices and an ethos of honesty, openness, and continuous reflection, including how research quality is judged" (Feuer, Towne & Shavelson, 2002b, p. 4, italics added). The question is whether AERA has such a culture, or instead has evolved a culture that is sometimes hostile to the exercise of judgment and criticism.

This is not to imply that I believe each of us is sufficiently expert to exercise judgment when confronting every kind of research. And we each should be sensitive to the reality that "When paradigms enter, as they must, into a debate about paradigmatic choice, their role is necessarily circular. Each group uses its own paradigm to argue in that paradigm's defense" (Kuhn, 1962, p. 93).

For me, the question of what is AERA's culture is delimited. The question is not about positivist versus nonpositivist paradigms. And I have never regarded quantitative methodology as inherently positivistic.

As Manicas and Secord (1983) noted, "All our perceptions,

categories, and frames of meaning are mediated and are culturally and historically loaded. But this does not eliminate the possibility of objectivity, construed here as warranted assertibility" (p. 410). Furthermore, quantitative researchers habitually do not in practice follow the procedural dicta of the quantitative "scientific method."

In practice, even for the most rigid quantitative researchers, the scientific method is a linear reporting sequence, and not a rubric for actual practice. Instead, science is distinguished by its bent of mind (Berliner, 2002). I tend to agree that the quantitative versus qualitative distinction in large part "is outmoded, and it does not map neatly in a one-to-one fashion onto any group or groupings of disciplines" (Shavelson & Towne, 2002, p. 19).

This does not mean that I fell expert enough to evaluate the rigor of warrants in every arena of educational research. But I do think the differences between paradigms have at times been drawn too starkly. As Howe and Eisenhart (1990) argued, "framing the issue of standards in terms of quantitative-qualitative debate is misguided" (p. 2).

Permission (Expectation) to Judge

I believe that when I attend sessions at the AERA annual meeting I should expect to see questioning (including self-questioning) regarding the warrants offered to support conclusions. But does the AERA culture or counter-culture deem such discussion politically incorrect?

In January, 1999, the AERA Council discussed the proposed theme for the 2000 annual meeting, "Defensible Knowledge Claims and

Diverse Voices: Making Theory and Methodology in the 21st Century".
The minutes of that meeting note that

> Dr. Shepard stated that although methodology is
> included in the title, the theme is not primarily
> about methodology. It is aimed at ways of knowing or
> developing theory. She noted that the theme is meant
> to draw attention to these ideas and to be inclusive
> as well as being critical. (AERA, 1999, p. 46)

The annual meeting theme ultimately was promulgated as "Creating Knowledge in the 21st Century: Insights from Multiple Perspectives."

During this era, AERA culture disapproved of even using the word, "methodology"; instead, the approved word was "warrants." In a similar vein, AERA editors were asked to no longer use the word "reject" in either review forms or in editorial action letters; instead, editors were asked to tell authors that the manuscript was "declined." At the October, 2007, meeting of the Mid-Western Educational Research Association (MWERA), one MWERA presenter who is an AERA editor noted that some authors receiving correspondence saying that the manuscript was "declined" were then contacting editors to ask if this meant that the manuscript was rejected!

Has it become impermissible to raise methodological issues, even if we defer to experts in given arenas of inquiry as to the appropriateness or thoughtfulness of different warrants? Shouldn't we instead be an organizational that engages in academic discourse and argument, without declaring selected topics taboo? What is the culture of AERA? What should our culture be?

Today, for some "...the very act of engaging the question

about the nature and quality of scientific research is, itself, suspect" (Feuer, Towne & Shavelson, 2002a, p. 28). Instead, the failure to question scientific rigor should be suspect. Methodology is not a bad word. In academic discourse we should expect to see experts in a given area of inquiry raising all kinds of questions about reasonableness.

At the Fall, 2002 AERA "Coordinated Committees" meeting, time was set aside for free-flowing discussion about what the annual meeting should look like in a decade. This naturally invoked some comments regarding what AERA is about. The single most important contribution that AERA makes to education is not the research that AERA helps members to disseminate.

I believe the single most important contribution that AERA makes to education is the creation and transmission of its culture. That culture frames the thinking and work of tens of thousands of educational researchers. It may be reasonable to create a culture that criticizes only criticism. But if that is the culture we decide to nurture, we ought to do so having thought carefully about the implications of our choices. I believe we would be best served by creating a culture that communicates a cherishing of the richness of diverse approaches to educational research, but also an expectation for rigor and an openness to dialogue about research quality.

Some Hopeful Signs

In 2006, AERA Council adopted its "Standards for Reporting on Empirical Social Science Research in AERA Publications" (AERA, 2006). The adoption of these standards thus mooted Fidler's (2002) lament regarding AERA's delinquency in addressing these issues. The

AERA (2006) standards noted that,

> ...[S]tatisticians have long warned against
> overreliance on significance testing to the
> exclusion of other methods of interpreting
> statistical analyses. Statistical significance tests
> combine both magnitude of relations (or estimates)
> and their uncertainty into the same quantity.
> Interpretation of statistical analyses is enhanced
> by reporting magnitude of relations (e.g., effect
> sizes) and their uncertainty separately... It is
> important to report the results of analyses that are
> critical for interpretation of findings in ways that
> capture the magnitude as well as the statistical
> significance of those results. (p. 37)

The milestones AERA's eventual movement toward effect size reporting include as benchmarks (Thompson, in press):

1988 First social science journal requires effect size reporting

1994 Second social science journal requires effect size reporting; the APA <u>Publication Manual</u>, used by more than 1,000 journals, first mentions effect sizes, and "encourages" their reporting

1996 APA appoints Task Force on Statistical Inference to make recommendations on whether statistical significance tests should be banned from APA journals

1999 Wilkinson and APA Task Force on Statistical Inference publish their recommendations in the <u>American Psychologist</u>, the APA flagship journal

2006 > 24 journals include explicit effect size reporting expectations within their author guidelines; AERA publishes its standards on reporting empirical research

This is not a portrait of courageous, forward-looking organizational leadership on issues of research rigor and integrity. Nevertheless, the movement away from a *laissez faire* refusal to articulate expectations involving research quality, as evidenced in the willingness to adopt and enforce standards within organizational publications, does bode well for the future. In my view, the willingness to promulgate standards--any standards--represents a sea change in AERA's organizational culture.

I hope the related dialogue will continue. Educational research would be strengthened by argument about standards. And AERA would be strengthened as an organization by continuing reflection on the organization's role in impacting educational research. It also remains my hope that other organizations, such as the National Academy of Education and the Spencer Foundation, will facilitate more organizational cultural naval gazing within AERA. We've got a lot to think about.

Room for Further Progress

Of course, there is always room for further improvement. As Levine (2007) noted in his recent report,

> As noted earlier, the American Educational Research
> Association (AERA), the largest research
> organization in education, has not served as an
> effective arbiter or monitor of quality. It has been
> unable to lead the profession in developing high,
> agreed-upon standards for quality research. Its

annual conference is more of a bazaar, displaying

the best and worst of education research. Indeed, **in**

interviews for this study, the deans of a number of

the highest-ranked graduate schools of education

lamented how much poor research is presented at

AERA. (p. 77, emphasis added)

Related comments were offered by Hess and LoGerfo (2006) in

"'Chicanas From Outer Space': Educational research is an

education!", which was their recent study of AERA conference

presentation titles (also see Eckardt, 2007).

Vital aspects of scholarship include exposing one's

conclusions and their warrants to public scrutiny, and

disseminating one's findings. Unfortunately, as against my early

experiences within AERA, today many presenters at the AERA annual

meeting do not seem to have prepared actual papers, and in other

cases presenters are unwilling to make available copies of their

works.

At a recent AERA Coordinated Committees meeting, it was

proposed that conference presenters should upload copies of their

papers into the All Academic system for access by interested AERA

members. All hell broke loose. Instead, today only presenters in

discussion sessions are **required** to upload copies of their

presentations in advance of the annual meeting, but these papers

are available **only** to session discussants. According to the 2008

Call for Proposals, participation in the new AERA Central

Repository of Annual Meeting Papers "by authors and presenters is

voluntary."

Certainly, there is room within the massive AERA annual

meeting program for people to make extemporaneous, unprepared remarks. But many authors who find themselves unable to write and disseminate work they are unwilling to stand behind might be encouraged to withhold their proposal submissions for a later year when their work may be more mature.

And the program would be more useful to members if the program publicly declared which presentations involve actual papers that will be deposited in the repository before the annual meeting. This would allow conference attendees to decide which sessions to go in cases in which these considerations are deemed relevant. As several AERA past presidents have noted, AERA stands for "American Educational Research Association," and not "American Educational Opinion Association" or "American Educational Discussion Association," though opinion and discussion both remain important even within an educational research association. But AERA members ought to be able walk into annual meeting sessions with some idea about what will be being served by the presenters.

References

American Educational Research Association. (1999). Council minutes. Educational Researcher, 28(3), 42-46.

American Educational Research Association. (2002). Research Advisory Committee. [retrieved from: **http://www.aera.net/whoswho/radcomm.htm**]

American Educational Research Association. (2006). Standards for reporting on empirical social science research in AERA publications. Educational Researcher, 35(6), 33-40.

American Psychological Association. (2001). Publication manual of the American Psychological Association (5th ed.). Washington, DC: Author.

Berliner, D.C. (2002). Educational research: The hardest science of all. Educational Researcher, 31(8), 18-20.

Capraro, R.M., & Thompson, B. (in press). What is an educational researcher, and what kinds of research will the educational researchers of tomorrow be trained to do? Journal of Educational Research.

Chambless, D. (1998). Defining empirically supported therapies. Journal of Consulting and Clinical Psychology, 66, 7-18.

Cook, C., & Heath, F. (2001). Users' perceptions of library service quality: A "LibQUAL+^{TM}" qualitative study. Library Trends, 49, 548-584.

Eckardt, N. (2007). The prevalence of qualitative methodology at AERA's annual meeting and the potential consequences. TCRecord. **http://www.tcrecord.org/Content.asp?ContentId=14741**

Eisner, E.W. (1983). Anastasia might still be alive, but the monarchy is dead. Educational Researcher, 12(5), 13-14, 23-24.

Fetterman, D.M. (1982). Ethnography in educational research: The dynamics of diffusion. Educational Researcher, 11(3), 17-22, 29.

Feuer, M.J., Towne, L., & Shavelson, R.J. (2002a). Reply. Educational Researcher, 31(8), 28-29.

Feuer, M.J., Towne, L., & Shavelson, R.J. (2002b). Scientific culture and educational research. Educational Researcher, 31(8), 4-14.

Fidler, F. (2002). The fifth edition of the APA Publication Manual: Why its statistics recommendations are so controversial. Educational and Psychological Measurement, 62, 749-770.

Fidler, F. (2005). From statistical significance to effect estimation: Statistical reform in psychology, medicine, and ecology. Doctoral dissertation, University of Melbourne. http://www.botany.unimelb.edu.au/envisci/docs/fidler/fidlerphd_aug06.pdf

Finch, S., Cumming, G., & Thomason, N. (2001). Reporting of statistical inference in the Journal of Applied Psychology: Little evidence of reform. Educational and Psychological Measurement, 61, 181-210.

Hess, F.M., & LoGerfo, L. (2006). "Chicanas From Outer Space": Educational research is an education! National Review Online. http://article.nationalreview.com/print/?q=ZDYwOGExMmUxOWYOZDgxNGQxMGEwZjg4NTNhMzQ2M2M=

Howe, K., & Eisenhart, M. (1990). Standards for qualitative (and quantitative) research: A prolegomenon. Educational Researcher, 19(4), 2-9.

International Committee of Medical Journal Editors. (1997). Uniform requirements for manuscripts submitted to biomedical journals. Journal of the American Medical Association, 277, 927-934.

Kuhn, T.S. (1962). The structure of scientific revolutions (2nd ed.). Chicago: University of Chicago Press.

Lagemann, E.C. (2000). An elusive science: The troubling history of education research. Chicago: University of Chicago Press.

Larabie, D.F. (2003). The peculiar problems of preparing educational researchers. Educational Researcher, 32(4), 13-22.

LeCompte, M.D., & Goetz, J.P. (1982). Problems of reliability and validity in ethnographic research. Review of Educational Research, 52, 31-60.

Levine, A. (2007). Educating researchers. Washington, DC: The Education Schools Project.

Manicas, P.T., & Secord, P.F. (1983). Implications for psychology of the new philosophy of science. American Psychologist, 38, 399-413.

Mosteller, F., & Boruch, R. (Eds.). (2002). Evidence matters: Randomized trials in education research. Washington, DC: Brookings Institution Press.

Robinson, D. (2004). An interview with Gene V Glass. Educational Researcher, 33(3), 26-30.

Sackett, D.L., Straus, S.E., Richardon, W. S., Rosenberg, W., & Haynes, R.B. (2000). Evidence-based medicine: How to practice and teach EBM (2nd ed.). New York: Churchill Livingstone.

Sanders, J.R. (1994). The process of developing national standards that meet ANSI guidelines. Journal of Experimental Education,

63, 5-12.

Sanders, J.R., & Joint Committee on Standards for Educational Evaluation. (1994). The program evaluation standards: How to assess evaluations of educational programs (2nd ed.). Newbury Park, CA: Sage.

Saxe, G., & Schoenfeld, A. (1998). Annual meeting 1999. Educational Researcher, 27(5), 41.

Shavelson, R.J., & Towne, L. (Eds.). (2002). Scientific research in education. Washington, DC: National Academy Press.

Shea, C. (1996). Psychologists debate accuracy of "significance test." Chronicle of Higher Education, 42(49), A12, A16.

Thompson, B. (2002). What future quantitative social science research could look like: Confidence intervals for effect sizes. Educational Researcher, 31(3), 24-31.

Thompson, B. (2006). Research synthesis: Effect sizes. In J. Green, G. Camilli, & P.B. Elmore (Eds.), Handbook of complementary methods in education research (pp. 583-603). Washington, DC: American Educational Research Association.

Thompson, B. (in press). Computing and interpreting effect sizes, confidence intervals, and confidence intervals for effect sizes. In J. Osborne (Ed.), Best practices in quantitative methods. Newbury Park, CA: Sage.

Thompson, B., Kyrillidou, M., & Cook, C. (in press). Library users' service desires: A LibQUAL+® study. Library Quarterly.

Towne, L., & Shavelson, R.J. (2003). Scientific research in education. Washington, DC: National Academies Press.

Towne, L., Wise, L.L., & Winters, T.M. (Eds.). (2005). Advancing

scientific research in education. Washington, DC: National Academies Press.

Wilkinson, L., & APA Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. American Psychologist, 54, 594-604.